Shanmukha Sai Dheeraz Chavali

+41783156414 | schavali@student.ethz.ch | LinkedIn | GitHub | Leetcode | HackerRank | Vaud, Switzerland

EDUCATION

ETH Zürich **Feb′2026 - Present**
Master of Science in Data Science (Special Student) Zürich, Switzerland
Relevant Coursework: Large Language Models, Scientific Computing, Data Modeling & Databases.

EPFL **Sept′2025 - Present**
Master of Science in Data Science (Exchange Year) (GPA: 5.1 / 6.0) Lausanne, Switzerland
Relevant Coursework: AI in Product Management, Modern NLP, Foundation Models and Gen AI, Graph Theory, Probabilistic models of modern AI, Ethics & Laws of AI, Applied Statistics, Data Intensive Systems.

University of Basel **Sept′2024 - Jul′2025**
Master of Science in Data Science (GPA: 5.0 / 6.0) Basel, Switzerland
Relevant Coursework: Mathematics for Data Science, Algorithms and Data Structures, Machine Learning, Artificial Intelligence, Reinforcement Learning, Machine Intelligence, Advanced Python.

SASTRA University **Jul′2017 - Jun′2021**
Bachelor of Technology in Electronics and Communication Engineering (GPA: 5.1 / 6.0) Thanjavur, India
Relevant Coursework: Advanced Calculus, Probability theory, Statistics & Numerical Methods, Python, Java and R.

PROFESSIONAL EXPERIENCE

Philip Morris International **Mar′2026 - Present**
Data Science and Artificial Intelligence Engineer Intern Vaud, Switzerland
- Architecting an enterprise-scale RAG pipeline on AWS Bedrock Knowledge Bases with semantic chunking and custom re-ranking, reducing hallucination rates by ~40% through contextual grounding guardrails.
- Building evaluation harnesses via SageMaker managed MLflow to A/B test 15+ prompt-model combinations across correctness, faithfulness, context relevance, and latency profiles.

Accenture **Feb′2022 - Aug’2024**
Advanced Application Engineering Sr. Analyst Bengaluru, India
- Designed and deployed a fully customized Amazon Connect instance integrated with Salesforce via Lambda, reducing third-party licensing dependency and saving the client **$40,000**/quarter in operational costs.
- Built an automated note-taking system for patient voice queries using Amazon Transcribe, Comprehend, S3 triggers, reducing manual documentation efforts and saving **$30,000**/quarter.
- Designed a predictive analytics pipeline over voice call logs and voicemails using S3, Glue, Redshift, Sage Maker, enabling real time concern clustering and saving **$60,000**/quarter through smarter call routing and resource allocation.

Indium Software **May′2021 - Dec′2021**
Junior Data Engineer Chennai, India
- Streamlined a hash-based skew mitigation algorithm, improving large join performance by 30% in fraud detection workloads.
- Refactored inefficient ETL jobs and restructured S3 storage classes with intelligent tiering, cutting monthly storage and compute expenses by $**25,000** while improving data retrieval SLAs.
- Designed hybrid vector-graph search system combining embedding-based semantic similarity with graph traversal algorithms, enabling complex multi-criteria customer segmentation with 85% precision improvement.

RESEARCH EXPERIENCE

University of Basel **Feb′2025 - Jan′2026**
Research and Teaching Assistant (Website) Basel, Switzerland
- Research Assistant under **Prof. Dr. Lukas Burkart**, spearheading interdisciplinary AI-humanities research to integrate LLMs and generative models into historical scholarship, transforming centuries-old texts into structured, machine-readable formats.
- Leveraging advanced Large Language Models (LLMs) to digitize and structure scholarly texts of ancient historians, improving online accessibility from data curation to model fine tuning and evaluation.
- Integrated graph-based RAG pipeline leveraging Neo4j and LangChain enabling multi-hop reasoning across complex narrative connections, improving context retrieval accuracy by 40% compared to traditional keyword-based search.

University of Basel **Nov'2024 - Feb'2025**

Research Assistant (Website) **Basel**, Switzerland

- Collaborated on the **SNSF-funded** Travail-Travel-Traders project, led by **Prof. Dr. Lesley Nicole Braun**, to analyze African women's trading practices and transnational connections between Africa and China.
- Developing scalable data pipelines to extract and process multimedia content from social media platforms using Beautiful Soup, Selenium, and asyncio. Implementing NLP models with Hugging Face Transformers and spaCy for sentiment analysis and topic modeling of social media content.
- Designing multimodal machine learning models using TensorFlow and PyTorch to analyze text, video, and metadata for cross-platform trade insights.

TECHNICAL SKILLS

Programming: Python (NumPy, SciPy, Pandas, Matplotlib, Seaborn), R, Java, C++, SQL and Linux.

Cloud & Big Data: AWS, Azure, Kafka, Big Query, Spark, Databricks, Airflow, Flink, Sage maker.

Machine Learning: Tensor Flow, Scikit-learn, XGBoost, CNN, RNN, LSTM, Deep learning, RAG, Pytorch, spaCy, Hugging Face.

Statistics: Statistical Inference, Bayesian Methods, Convex Optimization, Time-Series Analysis.

Miscellaneous: Power BI, Tableau, Databricks, DS & Algorithms, Docker, Linux, GitHub**,** Vector Search & Gen AI, MATLAB**.**

PROJECTS

Pedagogical Conversational Agents for Argumentation-based Learning

Tech Stack: *Python (Django), React, Vue.js, HTML, Machine Learning, REST API, Computer Vision*

- Designed and developed LLM-powered pedagogical conversational agents that integrate argument mapping visualization to scaffold students' critical thinking and argumentative reasoning skills in educational contexts.
- Addressed ethical AI considerations in educational technology by designing systems that prevent over-reliance on AI while promoting evidence-based reasoning, information literacy, and collaborative critical thinking skills.

Deep fake Detection

Tech Stack: *Python3, Deep learning, ResNeXt CNN, OpenCV, Computer Vision*

- Architected a hybrid CNN-LSTM pipeline combining ResNeXt-101 (spatial feature extraction) and bidirectional LSTMs (temporal analysis), achieving 89% accuracy on Face Forensics++ dataset (1K+ videos, 4 manipulation types).
- Improved robustness by retraining on emerging deepfake methods (StyleGAN3, Diffusion-based), sustaining 85% detection rates against adversarial attacks.
- Reduced false positives by 25% using k-fold cross-validation and attention mechanisms, outperforming baseline models.

Predictive Maintenance for Manufacturing Equipment

Teck Stack*: TensorFlow, PyTorch, Scikit-learn, LSTM, Temporal Fusion Transformer (TFT), XGBoost, Spark, AWS S3*

- Extended a predictive model to anticipate equipment failures in industrial settings using multivariate time-series sensor data.
- Crafted an ensemble model (XGBoost + Random Forest) with SMOTENC oversampling to handle imbalanced failure data.
- Reduced false alerts by 35% through PCA-driven feature selection.

Semantic Segmentation on Satellite Imagery

Tech Stack*: U-Net, Docker, Segmentation models, NumPy, Computer Vision, Matplotlib, Transformers*

- Orchestrated an end-to-end U-Net architecture with EfficientNet-B7 backbone, achieving 92% mIoU on LandCover.ai dataset (4 classes: buildings, water, woodland, background).
- Accelerated training by 40% using mixed-precision (FP16) and gradient checkpointing, optimizing for 5120x5120px satellite images.
- Integrated Weights & Biases for experiment tracking, reducing hyperparameter tuning cycles by 30% through automated LR schedule optimization.

ACHIEVEMENTS AND POSITIONS OF RESPONSIBILITY

- Member of IT - EPFL Rocket Team
- Head of employee engagement activities.
- Workshop Curator - IoT & Robotics Outreach
- Winner of El-com Dias | Intra-University project competitions.
- Captain - State level Volleyball team.
- Winner of the district-level chess championship.